Inverse Rendering for Discrete X-Ray Computed Tomography

LOVRO NUIC, École Polytechnique Fédérale de Lausanne (EPFL), Switzerland
ZIYI ZHANG, École Polytechnique Fédérale de Lausanne (EPFL), Switzerland
KORBINIAN SAGER, Carl Zeiss AG, Germany
WENZEL JAKOB, École Polytechnique Fédérale de Lausanne (EPFL), Switzerland

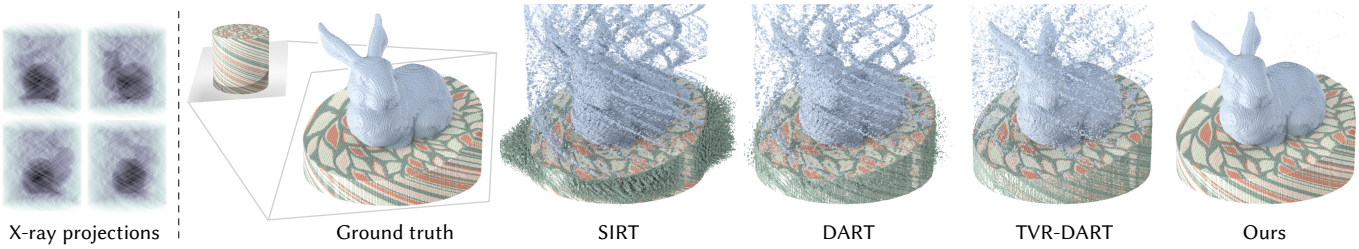


Fig. 1. Cutaway views of reconstructions of a cylindrical sample composed of three materials with known X-ray absorption coefficients. Standard methods compute the error in X-ray projection space and back-project this error, potentially incorporating constraints or regularization to promote discrete solutions. Our method instead minimizes a volumetric loss function that considers conflicting alternative material assignments. It consistently produces superior reconstructions, particularly in challenging cases with sparse or noisy projections.

Discrete X-ray tomography reconstructs the internal structure of an object from X-ray projections, assuming that the volume is composed of a discrete set of known materials (e.g., steel, aluminum, and air). This is generally straightforward when many projections are available but becomes increasingly ill-posed as their number decreases. Discrete tomography has been extensively studied over the past five decades, resulting in a range of mature reconstruction algorithms.

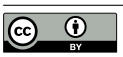
In this work, we introduce a new reconstruction method that draws inspiration from both classical computed tomography and recent advances in inverse rendering, demonstrating that a remarkably simple gradient-based inversion can significantly surpass the reconstruction quality of standard methods such as SIRT, DART, and TVR-DART. Our method represents each 3D location as a probability distribution over the set of known materials and minimizes a volumetric loss that encourages consistency with the measured projections. It supports nonlinear effects such as volumetric scattering and is simple to optimize and parallelize on compute accelerators.

We evaluate our method on challenging 2D and 3D benchmarks, demonstrating superior performance particularly in sparse and limited-angle scenarios, where traditional techniques struggle with ambiguity.

CCS Concepts: • **Computing methodologies** → **Reconstruction**; *Rendering*; *3D imaging*.

ACM Reference Format:
Lovro Nuic, Ziyi Zhang, Korbinian Sager, and Wenzel Jakob. 2026. Inverse Rendering for Discrete X-Ray Computed Tomography. *ACM Trans. Graph.* 45, 4, Article 134 (July 2026), 16 pages. https://doi.org/10.1145/3811391

Authors' Contact Information: Lovro Nuic, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, lovro.nuic@epfl.ch; Ziyi Zhang, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, ziyi.zhang@epfl.ch; Korbinian Sager, Carl Zeiss AG, Oberkochen, Germany; Wenzel Jakob, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, wenzel.jakob@epfl.ch.



ACM 1557-7368/2026/7-ART134
https://doi.org/10.1145/3811391

1 Introduction

Computed tomography (CT) reveals an object's internal structure by analyzing its interaction with a given form of energy. X-ray CT, which relies on wavelengths between 0.01 and 10 nanometers, remains the most widely used approach due to its strong penetrating ability, high sensitivity to material differences, and comparatively low acquisition cost. Other energy forms lead to tomography methods such as magnetic resonance imaging (MRI), positron emission tomography (PET), and electron tomography (ET).

This work focuses on X-ray computed tomography[1], where a two-dimensional detector array captures radiation emitted by an X-ray source. An object placed between the source and the detector attenuates a portion of this radiation, revealing information about the sample's atomic composition and density. By collecting projections from multiple orientations—typically by rotating the object or the source-detector system—the inverse problem of reconstructing the object's internal structure becomes well-posed.

In practice, however, acquiring high-resolution projections that densely cover the full 360° range is often infeasible due to constraints on acquisition time, data storage, mechanics (e.g., restricted rotation range), and concerns about exposing specimens to large amounts of ionizing radiation (e.g., when imaging organic material). Fortunately, the number of projections required for an accurate reconstruction can be significantly reduced by leveraging prior knowledge, such as *discrete tomography*, which models the volume as a composition of a finite set of known materials rather than a continuous density. This dramatically reduces the size of the solution space—for example, a reconstruction of an object composed of four materials only requires 2 bits of information per cell, which explains why discrete tomography can achieve substantially better quality with fewer and lower-quality projections.

[1]Our method may also be applicable to CT modalities based on similar mathematical principles (e.g., PET, ET), but such generalizations are beyond the scope of this work.

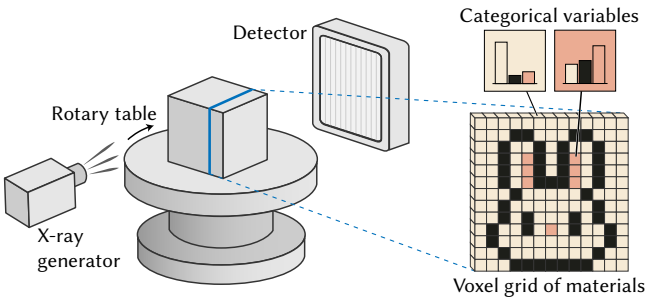


Fig. 2. **Schematic of our discrete CT model.** X-rays pass through an object represented by a voxel grid of materials, where they undergo attenuation and scattering, and then reach a detector on the right. The input consists of multiple projections captured from different orientations (e.g., using a rotation stage). Each voxel encodes a categorical distribution over materials, which we optimize to produce a single material assignment.

Previous approaches to discrete CT have largely repurposed continuous methods by adding regularization and constraints that promote discrete solutions. Our contribution is a new reconstruction method that directly addresses the discrete problem without relying on such heuristics.

Our work is motivated by recent advances in inverse rendering for geometry reconstruction. These methods typically refine a 3D shape by minimizing the difference between a rendered image and a reference—a process analogous to CT reconstruction, where rendering corresponds to the Radon transform. Like in discrete CT, many inverse rendering approaches first optimize a continuous representation (e.g., a volumetric field) and then extract discrete geometry from it. Recently, two works in graphics [Zhang et al. 2024, 2025] pursued an alternative that deviates from the standard practice of differentiating an image formation model to obtain gradients: they optimize a probabilistic superposition that models competing alternative explanations of the observed data. As the optimization progresses, this distribution naturally collapses to a single geometric solution.

While these two prior surface reconstruction methods do not generalize to discrete CT, they motivate our approach based on a probabilistic model of discrete material assignments (Figure 2). This produces significantly sharper and more accurate boundaries than prior techniques because its objective function drives the optimization to converge to a single material label at each position. Additional material-specific effects like polychromatic absorption or scattering can be naturally incorporated while retaining this fundamental behavior.

We implemented our method in a physically based inverse rendering framework and demonstrate its generality and effectiveness across a wide range of tasks, finding that it consistently outperforms existing methods both in terms of reconstruction quality and robustness to sparse and limited-angle configurations.

2 Related Work

We now review the literature on continuous and discrete CT along with relevant developments in the areas of graphics and vision.

2.1 Continuous reconstruction

Filtered Backprojection (FBP) [Bracewell and Riddle 1967; Cormack 1964; Ramachandran and Lakshminarayanan 1971] is a foundational analytic reconstruction technique that realizes the inverse of the Radon [1917] transform. Given perfect data, FBP and its extensions [Feldkamp et al. 1984] reliably produce reconstructions of outstanding quality with a computational efficiency that makes them suitable for routine clinical and industrial applications.

However, when data suffers from inadequate angular sampling, restricted field of view, or measurement noise, analytic methods tend to produce low-quality reconstructions contaminated with substantial artifacts.

Iterative reconstruction methods like the Simultaneous Algebraic Reconstruction Technique (SART) [Andersen and Kak 1984], Simultaneous Iterative Reconstruction Technique (SIRT) [Gilbert 1972] and Maximum Likelihood Expectation Maximization (MLEM) [Shepp and Vardi 1983] are preferable in such cases. They incorporate statistical models, regularization strategies, and prior knowledge to refine a tentative solution to improve quality and robustness.

2.2 Discrete reconstruction

The Discrete Algebraic Reconstruction Technique (DART) [Batenburg and Sijbers 2011] introduces an efficient heuristic for discrete tomography by alternating between hard-threshold segmentation and continuous algebraic updates. Assuming high reconstruction accuracy in homogeneous regions, DART focuses the algebraic refinement on boundaries.

Several variants of DART further improve computational efficiency [Dabravolski et al. 2014; Frenkel et al. 2023; Maestre-Deusto et al. 2011], noise handling [Bleichrodt et al. 2014], and additionally determine unknown gray-levels (i.e., absorption coefficients) and associated segmentation thresholds [He et al. 2022; Huang et al. 2023; van Aarle et al. 2012]. Poly-DART [Six et al. 2019] handles beam-hardening in polychromatic X-ray data.

2.3 Non-convex methods

The works listed above all incorporate continuous steps that require solving linear systems of equations. Some prior works also adopt non-convex objectives, e.g., to smoothly guide a continuous solution toward a discrete material assignment [Rashed and Kudo 2012]. TVR-DART [Zhuge et al. 2016] builds on this idea and additionally imposes a total variation prior to encourage sparse material transitions.

Most X-ray CT reconstruction methods assign a scalar X-ray absorption value to each voxel. The work of Zisler et al. [2016] is notable for adopting a categorical representation, assigning a material class membership vector to each voxel. Our method also uses a categorical representation and imposes a total variation prior, but it does so within a volumetric optimization framework that differs substantially from projection-space error minimization.

We selected SIRT, DART, and TVR-DART as representatives of the classes of methods covered in the previous three subsections and use them as experimental baselines. We also compare to the method of Zisler et al. [2016].

2.4 Field representations in graphics and vision

The fields of computer graphics and computer vision have seen substantial research on emissive field representations for novel view synthesis, such as NeRF [Mildenhall et al. 2020] and Gaussian Splats [Kerbl et al. 2023]. The underlying ideas are portable to the X-ray CT setting and can likewise be used to represent 3D density, which not only compresses the data but also provides an effective regularizer that stabilizes the reconstruction. Work in this vein includes reconstruction methods using *multi-layer perceptron* (MLP)-based field representations [Sun et al. 2021; Zang et al. 2021], MLPs augmented with hierarchical feature grids [Rückert et al. 2022], and Gaussian splats [Wang et al. 2025; Zha et al. 2024]. It is also possible to drop the representation altogether and define a grid-free random walk, whose 3D sample density implicitly defines the desired volume [Gregson et al. 2012].

We use a relatively ordinary 3D grid representation that stores a discrete probability vector per voxel. This representation could also be encoded into a neural field, though we choose not to do so for simplicity. Our main contribution lies in how we optimize this representation to determine per-voxel membership in one of multiple material classes. Our method builds on a physically based renderer and can also account for interreflection, which these previous methods lack.

2.5 Machine learning

A large body of prior work covers learning-based denoising, artifact removal, and segmentation methods that are often combined with existing X-ray CT reconstruction algorithms. While such steps could potentially be incorporated into our method, we consider them out of scope for this submission, which presents a classical solution that does not require training on prior datasets.

2.6 Quality metrics

Given a tentative reconstruction, it would be desirable to inspect a 3D error map to identify inaccurately reconstructed regions. Since ground-truth errors are typically unknown, previous work has introduced proxy metrics. Notably, Roelandts et al. [2014] introduced the Reconstructed Residual Error (RRE), obtained by back-projecting the difference between computed and measured projections, and showed that it strongly correlates with true reconstruction error in synthetic and real-world CT tasks. The RRE was originally conceived as a post-reconstruction diagnostic. Our method can be interpreted as a new optimization algorithm that directly minimizes the expectation of residual errors in the RRE domain.

3 Method

The main contribution of this paper broadly applies to inverse rendering with discrete parameter spaces. We therefore first introduce the idea at a high level and then specialize it to CT.

Consider minimizing a rendering loss $\mathcal{L}(\boldsymbol{X})$ with respect to variables $\boldsymbol{X} = (X_1, \ldots, X_N) \in \Omega$ representing the scene. Standard methods like *path replay backpropagation* [Vicini et al. 2021] apply when Ω is continuous but fail for discrete Ω, where $\mathcal{L}$ is not differentiable.

A common strategy is to relax the discrete formulation into a continuous one, for example by replacing discrete material assignments with convex combinations of reference materials [Gkioulekas et al. 2013; Nindel et al. 2021]. Gradient-based methods can then optimize the mixing weights, though this generally leads to material blends. Although fractional assignments may fit observations well, they are not valid solutions in our setting. Rounding the recovered parameters to discrete assignments can in turn severely degrade quality because this added step was not considered in the optimization. While we also use a continuous relaxation to benefit from gradient-based methods, we design it to be sensitive to the discrete nature of the parameter space to avoid this drop in quality.

3.1 Probabilistic formulation

We model the assignment $\boldsymbol{X}$ as a discrete random variable with mass function $p(\boldsymbol{X} \mid \boldsymbol{\Theta})$, where $\boldsymbol{\Theta}$ denotes a set of distribution parameters. Minimizing the expected loss under this distribution yields an alternative relaxation that optimizes $\boldsymbol{\Theta}$ in a continuous space, while only evaluating $\mathcal{L}(\boldsymbol{X})$ at valid discrete configurations:

$$\begin{aligned}\boldsymbol{\Theta}^* &= \arg\min_{\boldsymbol{\Theta}} \; \mathbb{E}_{\boldsymbol{X}\sim p(\boldsymbol{X}\mid\boldsymbol{\Theta})}\left[\mathcal{L}(\boldsymbol{X})\right] \\ &= \arg\min_{\boldsymbol{\Theta}} \; \sum_{\boldsymbol{X}\in\Omega} p(\boldsymbol{X} \mid \boldsymbol{\Theta})\, \mathcal{L}(\boldsymbol{X}). \end{aligned} \tag{1}$$

Unlike mixture-based relaxations that find the most plausible blend, this formulation models our belief about the underlying discrete configuration given the observed data.

Equation 1 is impractical as stated: for N variables with m discrete states, the distribution assigns probability mass to m^N possible configurations, which is far too large to store or explore during optimization. We therefore restrict it to a factorized family of independent categorical distributions:

$$p(\boldsymbol{X} \mid \boldsymbol{\Theta}) = \prod_{i=1}^{N} p(X_i \mid \boldsymbol{\theta}_i), \qquad X_i \sim \mathrm{Cat}(\boldsymbol{\theta}_i), \tag{2}$$

where the vectors $\boldsymbol{\theta}_i = (\theta_{i,1}, \ldots, \theta_{i,m})$ lie on the probability simplex $\Delta_{m-1} = \{\boldsymbol{\theta} \in \mathbb{R}^m \mid \theta_k \geq 0,\ \sum_{k=1}^{m} \theta_k = 1\}$. The family remains expressive[2]: it contains all deterministic assignments as special cases (when each $\boldsymbol{\theta}_i$ is one-hot), and gradient-based optimization can navigate between them using the continuous simplex domain.

The factorization in Equation (2) is an approximation, not a claim that the variables are truly independent. The rendering loss already couples them, and we will later augment the objective with spatial regularization to suppress spurious small-scale variation in the assignment field. This couples neighboring variables and reintroduces spatial dependencies in the otherwise factorized model (2). Equivalently, this regularization can be interpreted as a prior in a Bayesian formulation.

The principles discussed so far are not new: mean-field factorizations are common in variational inference [Bishop and Nasrabadi 2006] and have also been used for discrete optimization in Bayesian

[2]The factorized family cannot represent distributions with correlated variables. When multiple distinct solutions achieve equally low loss, the optimization may fail to select one and assign intermediate probabilities. We observe reliable convergence to deterministic assignments when the inverse problem has a unique or dominant solution.

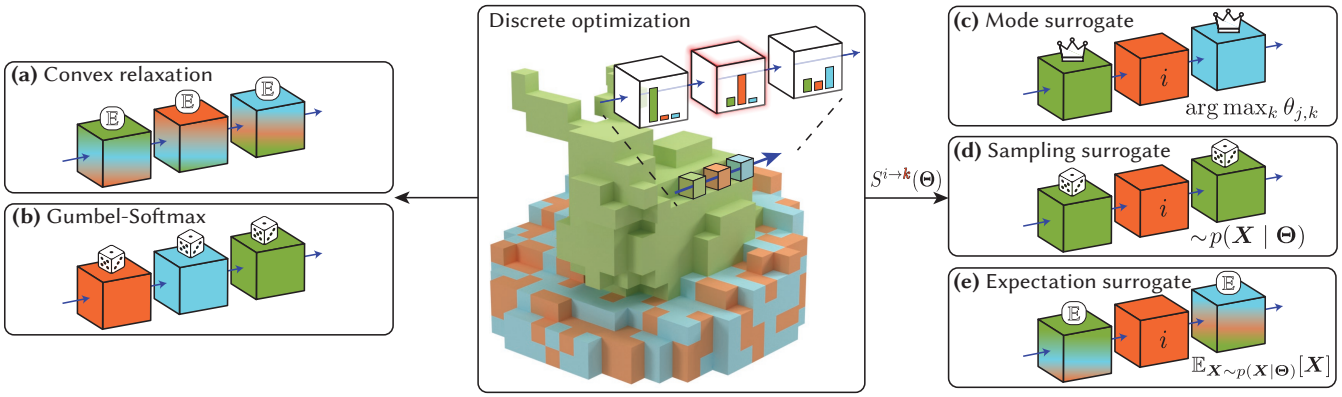


Fig. 3. **Categorical optimization.** We model each voxel's material assignment using a categorically distributed random variable, allowing the optimization to gradually evolve its confidence in this discrete choice. The figure shows a projection ray passing through three voxels, comparing different formulations that make this problem differentiable: (**a**) convex relaxation optimizes the expected (blended) material; (**b**) Gumbel-Softmax differentiably reparameterizes the random sampling process; (**c**–**e**) our surrogates collapse the distribution to a single state (via its mode, a random sample, or expectation) and compute gradients by evaluating the loss of voxel i under each possible material assignment k.

settings [Daulton et al. 2022]. We build on this foundation to systematically explore a large design space of estimators and update rules to assess their potential for inverse rendering.

3.2 Ideal gradient

We use the law of total expectation to reformulate the expected loss and capture the influence of a specific assignment X_i:

$$\begin{aligned}\mathbb{E}[\mathcal{L}(\boldsymbol{X})] &= \mathbb{E}[\mathbb{E}[\mathcal{L}(\boldsymbol{X})|X_i]] \\ &= \sum_{k=1}^{m} \theta_{i,k} \underbrace{\mathbb{E}[\mathcal{L}(\boldsymbol{X})|X_i = k]}_{=\mathcal{L}_i(k)}. \end{aligned} \tag{3}$$

This equation holds for any $i = 1 \dots N$. The *conditional loss* function $\mathcal{L}_i(k)$ represents the expected loss when variable X_i is forced to state k. Differentiating Equation (3) with respect to an arbitrary parameter $\theta_{i,k}$ yields:

$$\frac{\partial}{\partial \theta_{i,k}} \mathbb{E}[\mathcal{L}(\boldsymbol{X})] = \mathcal{L}_i(k). \tag{4}$$

We refer to this as the **ideal gradient**. It describes how increasing the probability mass of state k changes the objective in proportion to the conditional loss $\mathcal{L}_i(k)$. Taking a step along the negative gradient therefore shifts probability mass toward lower-loss candidates. The compact form of this expression follows directly from the factored probability model defined in Equation 2.

This gradient is unconstrained in the sense that it does not consider the geometry of the simplex domain Δ_{m-1} (increasing $\theta_{i,k}$ requires decreasing one of the $\theta_{i,k'}$). We will revisit this point later.

3.3 Surrogate approximation

We previously switched to a factored model $p(\boldsymbol{X} \mid \boldsymbol{\Theta})$ to avoid having to track an exponential number of discrete probabilities. Unfortunately, the conditional loss $\mathcal{L}_i$ still requires marginalizing over m^{N-1} remaining configurations, which is intractable.

To make optimization practical, we introduce a *surrogate map* $\mathcal{S} : \Delta_{m-1}^N \to \Omega$ that collapses the distribution to a single configuration. We further write $\mathcal{S}^{i\to k}(\boldsymbol{\Theta})$ for the configuration obtained by first forcing variable i to state k and then applying the surrogate to the remaining variables. This yields an approximate conditional loss:

$$\tilde{\mathcal{L}}_i(k) \coloneqq \mathcal{L}(\mathcal{S}^{i\to k}(\boldsymbol{\Theta})) \approx \mathcal{L}_i(k). \tag{5}$$

We consider three different surrogates:

- **Mode.** Pick the most probable state, i.e., $\mathcal{S}(\boldsymbol{\Theta})_j = \arg\max_k \theta_{j,k}$. This ignores uncertainty before the distribution concentrates.
- **Sample.** Draw $\mathcal{S}(\boldsymbol{\Theta}) \sim p(\boldsymbol{X} \mid \boldsymbol{\Theta})$ and average the loss over one or more samples. This is unbiased but may exhibit high variance.
- **Expectation.** Define $\mathcal{S}(\boldsymbol{\Theta}) = \mathbb{E}_{\boldsymbol{X}\sim p(\boldsymbol{X}|\boldsymbol{\Theta})}[\boldsymbol{X}]$, where the expectation is taken over material properties. This is generally an approximation, as $\mathcal{L}(\mathbb{E}[\boldsymbol{X}]) \neq \mathbb{E}[\mathcal{L}(\boldsymbol{X})]$. While this surrogate produces blended materials, the resulting gradient formulation retains its discrete character by evaluating how switching each variable to a specific state would affect the loss.

The expectation-based surrogate performs best in our experience. Unlike the mode, it incorporates the full distribution; unlike sampling, it introduces no variance.

3.4 Surrogate gradient

The introduction of surrogates leads to an alternative gradient formulation. Recall how Equation 3 introduced a local decomposition of the objective $\mathbb{E}[\mathcal{L}(\boldsymbol{X})]$ conditioned on a single variable. This equation holds for every $i \in \{1, \dots, N\}$, meaning that the same objective can be written in N equivalent ways.

Collapsing the distribution with a surrogate map breaks this equivalence: the resulting *approximate* marginalization for each variable i now provides a different view of how assignments affect the loss:

$$\tilde{\mathcal{L}}^{(i)}(\boldsymbol{\Theta}) = \sum_{k=1}^{m} \theta_{i,k} \tilde{\mathcal{L}}_i(k). \tag{6}$$

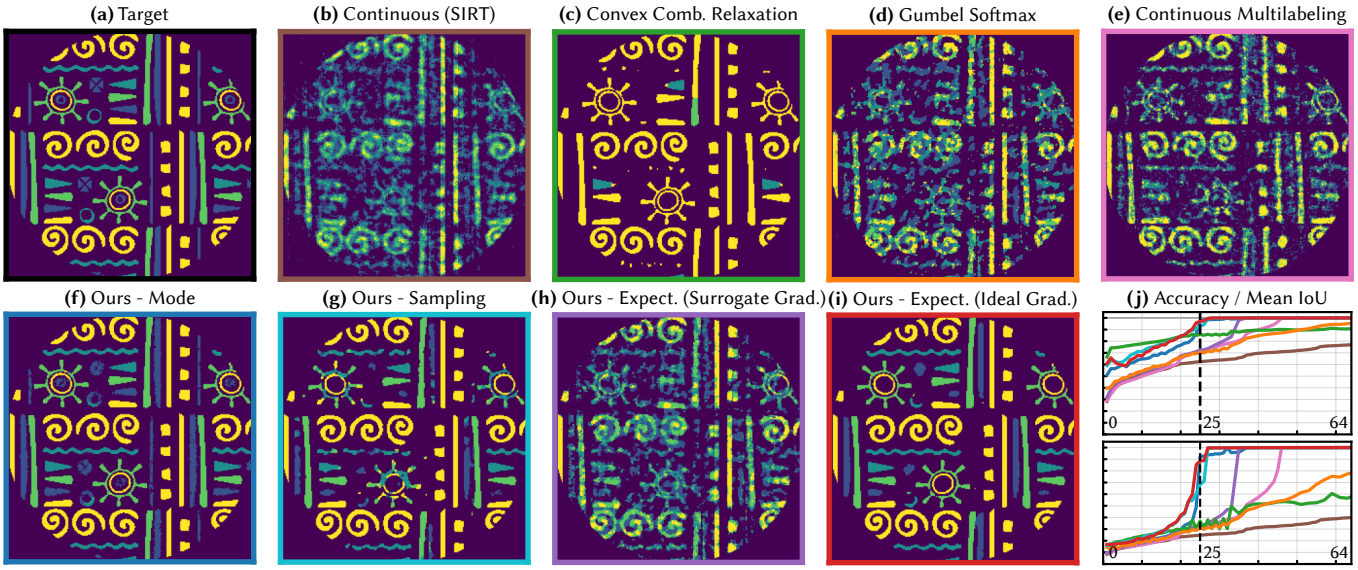


Fig. 4. **Problem formulations.** We compare the effectiveness of different continuous and probabilistic formulations in a synthetic 5-material CT reconstruction: (**a**) reference (2D), (**b**) continuous optimization of extinction with rounding [Gilbert 1972], (**c**) convex combinations of reference materials with a discreteness-promoting entropy loss [Gkioulekas et al. 2013; Nindel et al. 2021], (**d**) Gumbel-Softmax with a tuned temperature schedule, (**e**) continuous multilabeling [Zisler et al. 2016], **(f–i)** four variants of the proposed method using mode, sampling, and expectation-based surrogates; the latter further admits two gradient formulations (surrogate and ideal). Plot **(j)** shows reconstruction accuracy (% of correct assignments) and IoU (intersection-over-union, i.e., the average per-class overlap with ground truth) as a function of projection count. The shown reconstructions are for 25 projections (dashed vertical line). We find that the expectation-based surrogate with ideal gradients in (**i**) produces the fastest and most reliable convergence. The trends shown here also generalize to more challenging 3D reconstructions.

To treat all variables symmetrically, we define the surrogate objective $\mathcal{J}(\boldsymbol{\Theta}) \approx \mathbb{E}[\mathcal{L}(\boldsymbol{X})]$ as their average:

$$\mathcal{J}(\boldsymbol{\Theta}) = \frac{1}{N}\sum_{i=1}^{N} \tilde{\mathcal{L}}^{(i)}(\boldsymbol{\Theta}) = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{m} \theta_{i,k}\tilde{\mathcal{L}}_i(k). \tag{7}$$

Differentiating with respect to $\theta_{i,k}$ yields the **surrogate gradient**:

$$\frac{\partial}{\partial\theta_{i,k}}\mathcal{J}(\boldsymbol{\Theta}) = \frac{1}{N}\tilde{\mathcal{L}}_i(k) + \frac{1}{N}\sum_{j\neq i}\sum_{\ell=1}^{m}\theta_{j,\ell}\frac{\partial\tilde{\mathcal{L}}_j(\ell)}{\partial\theta_{i,k}}. \tag{8}$$

The first term captures the direct influence of $\theta_{i,k}$ through its own marginalization, analogous to the ideal gradient. The second term represents coupling between coordinate views: how changes in $\theta_{i,k}$ affect the surrogate evaluations in other marginalizations. For the mode and sample surrogates, this coupling term vanishes as their evaluations do not continuously depend on $\theta_{i,k}$ for $j \neq i$. For the expectation surrogate, the term is nonzero and can be computed via backpropagation through the renderer.

We can also adapt the **ideal gradient** from Equation 4 by replacing the intractable conditional loss $\mathcal{L}_i(k)$ with its approximation:

$$\frac{\partial}{\partial\theta_{i,k}}\mathcal{J}(\boldsymbol{\Theta}) \approx \tilde{\mathcal{L}}_i(k), \tag{9}$$

which lacks the coupling terms and is therefore easier to compute.

Although the surrogate gradient is technically the correct derivative of $\mathcal{J}$, we find that the ideal gradient motivated by the exact theory leads to more stable convergence in practice.

Gumbel–Softmax. We also investigated another common differentiable relaxation of categorical distribution sampling known as the Gumbel–Softmax approach [Jang et al. 2017]. This method does not leverage the previously introduced framework and instead replaces each discrete assignment X_i with a continuously parameterized vector $\boldsymbol{X_i}$ defined as

$$\boldsymbol{X}_i = \mathrm{softmax}((\log\boldsymbol{\theta}_i + \boldsymbol{G}_i)/\tau) \tag{10}$$

where $G_{i,k} \sim \mathrm{G}(0,1)$ are i.i.d. samples from the Gumbel distribution and τ is a temperature parameter. As $\tau \to 0$, the probability mass converges on a single entry whose index follows $\mathrm{Cat}(\boldsymbol{\theta}_i)$.

This reparameterization works best with a temperature schedule that starts high to enable exploration and then gradually decreases throughout optimization to ensure convergence to a discrete assignment. Even with tuning, we found that this approach converges substantially more slowly and produces lower quality results, with the gap widening as the problem size (number of voxels) increases.

3.5 Design space exploration (i)

While we have not yet discussed CT (Section 3.8) in detail, it is useful to already explore the impact of these choices on a concrete problem. Figure 4 compares continuous and probabilistic formulations on a synthetic 2D phantom, which becomes easier to reconstruct as the number of input projections increases (Figure 4j). We briefly summarize key findings here; Section 2 in the supplemental material provides additional discussion and details on each experiment.

The common failure mode of baseline methods (**b**-**e**) is that gradients pull the optimization toward blended configurations. SIRT then incurs rounding errors; convex relaxation mostly interpolates between extreme materials; bias in the Gumbel-Softmax parameterization limits quality; continuous multilabeling is sensitive to the balance between data fidelity and discretization.

Among our surrogates, the mode approach prematurely commits to a discrete class when uncertainty is high early in the optimization. The sample surrogate avoids this but introduces gradient variance that slows convergence. The expectation surrogate accounts for the full distribution without adding variance and converges most reliably. Of its two gradient formulations, the ideal gradient (Eq.9) outperforms the full surrogate gradient (Eq.8) because the coupling terms in the latter introduce continuous objectives that compete with discrete convergence.

Discussion. Standard practice in inverse rendering is to derive physically based reconstruction objectives by differentiating an image formation model with respect to its parameters. Rather than proposing a new forward model, we directly focus on objectives and gradients that enable effective optimization in discrete spaces. Our approach considers a large set of discrete moves around the current state. This systematic probing of the loss landscape facilitates exploration of the combinatorial assignment spaces, in ways that are difficult to achieve with existing continuous relaxation schemes.

3.6 Simplex constraint

The previous discussion neglected the non-Euclidean geometry of the simplex domain Δ_{m-1}. We experimented with four methods to keep the parameters $\boldsymbol{\theta}_i$ on the probability simplex.

- **Softmax parameterization.** We introduce unconstrained logits $\boldsymbol{\ell}_i \in \mathbb{R}^m$ and set $\theta_{i,k} = \exp(\ell_{i,k})/\sum_{k'} \exp(\ell_{i,k'})$, which satisfies the simplex constraint by construction. The downside is saturation: when logits reach extreme values, probabilities saturate and resist further change, which slows convergence. Note that this option is not related to the Gumbel–Softmax approach discussed earlier.
- **Projected gradient descent.** We can combine the unconstrained gradients with the method of Wang and Carreira-Perpiñán [2013] to project the optimizer state onto the simplex following each step. This avoids saturation but introduces another problem: the two gradient formulations (4, 8) blend losses and are inherently positive prior to projection. The efficiency of adaptive optimizers like Adam is in part a result of tracking sign variation in moment accumulators, and we found them to perform poorly when consistently driven by all-positive gradients.
- **Natural gradient**. The Euclidean gradient ignores the simplex geometry and is not the steepest-ascent direction under the manifold's metric. We can instead use the *natural gradient* $\hat{\mathbf{g}}$ [Amari 2016], which on the simplex admits the closed-form expression

$$\hat{\mathbf{g}} = \boldsymbol{\theta} \odot (\mathbf{g} - \langle \mathbf{g}, \boldsymbol{\theta} \rangle \mathbf{1}), \tag{11}$$

  where $\odot$ denotes component-wise multiplication with the probability vector $\boldsymbol{\theta}$ at which the gradient was evaluated. Natural gradient descent alone does not guarantee that the state remains on the manifold and must still be combined with a projection.

- **Modified natural gradient**. Equation 11 reveals a practical drawback: scaling by $\boldsymbol{\theta}$ impedes steps into low-probability directions, which we found to slow convergence. We define a modified natural gradient by removing the component-wise scaling, which yields

$$\hat{\mathbf{g}} = \mathbf{g} - \langle \mathbf{g}, \boldsymbol{\theta} \rangle \mathbf{1}, \tag{12}$$

  and combine it with the simplex projection.

3.7 Design space exploration (ii)

Figure 5 compares the four simplex constraint strategies across three optimizers (SGD with/without momentum and Adam), using the phantom from Figure 4a with 32 projections.

The softmax parameterization and natural gradient both exhibit slow convergence because they make the corners of the probability simplex "sticky," which impedes discrete flips that are needed to change material assignments during the optimization. Projected gradient descent shows modest performance with SGD but fails with Adam, where the all-positive entries of the ideal gradient interfere with the method's adaptive step size computation.

Our modified natural gradient formulation resolves this by subtracting the expected loss under the current distribution from each component, so that the resulting gradient records how much better or worse each material assignment is relative to this baseline. These naturally signed values restore the effectiveness of step size adaptation. Under SGD, this update coincides with projected gradient descent, because the two directions differ only by a constant shift applied equally to every component, which the simplex projection cancels out. In our experiments, modified natural gradient with Adam converges roughly an order of magnitude faster than other combinations, so we adopt it for all remaining experiments.

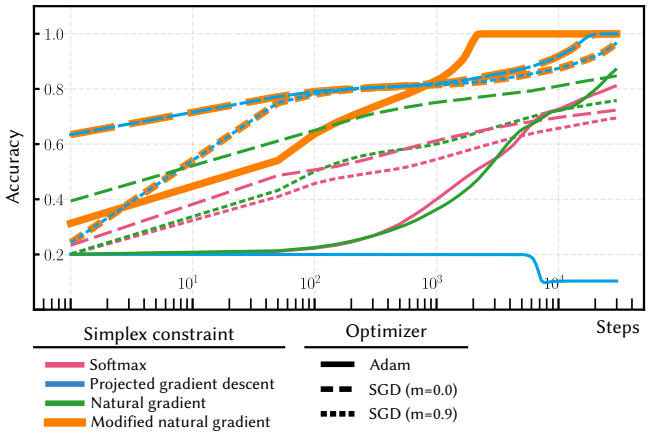


Fig. 5. **Constraint formulations.** We compare four strategies to keep parameters $\boldsymbol{\theta}_i$ on the probability simplex (so that $\sum_k \boldsymbol{\theta}_{i,k} = 1$): a softmax parameterization (which satisfies the constraint by construction), projected gradient descent, projected natural gradients (which better respect the non-Euclidean geometry of the simplex), and our proposed modified natural gradient. The modified natural gradient with Adam converges roughly an order of magnitude faster than alternatives.

Convergence behavior. When optimizing using ideal gradients and any of the three surrogates, the distributions $\boldsymbol{\theta}$ naturally concentrate as the optimization progresses: probability mass shifts toward one candidate per variable, approaching a deterministic assignment. This behavior emerges from the structure of the objective rather than being explicitly enforced. To see why, note that each per-variable loss $\mathcal{J}_i(\boldsymbol{\theta}_i) = \sum_k \theta_{i,k} \tilde{\mathcal{L}}_i(k)$ is a convex combination of the candidate losses. Since this is a linear function over the probability simplex, its minimum occurs at a vertex (a one-hot vector) unless multiple candidates achieve equal loss, in which case any mixture of tied candidates is also optimal.

When the loss landscape has a clear minimum, gradient descent drives the distributions toward these vertices. In Section 1 of the supplemental material, we prove that the optimization converges to a *1-neighbor local minimum*: a deterministic configuration where no single-variable change can reduce the loss.

This convergence property does not imply that the optimization finds a global minimum, as many discrete inverse problems are NP-hard [Gardner et al. 1999; Gerard 2009]. However, it ensures that the final assignment is locally optimal with respect to individual variable changes, which in our experiments yields high-quality solutions.

3.8 Discrete Computed Tomography

We now specialize our modeling framework to discrete CT, focusing first on the absorptive case without scattering. Index i refers to the i-th voxel, and the associated variable $X_i \in \{1, \dots, m\}$ assigns one of m possible discrete materials. For a ray $\mathbf{r}(t) = \mathbf{o} + t\boldsymbol{\omega}$, the Beer–Lambert law gives the transmitted intensity as

$$I(\boldsymbol{X}, \mathbf{r}) = I_0 \exp\left(-\sum_{i \in \mathbf{r}} \mu(X_i)\, \Delta_i\right), \tag{13}$$

where "$i \in \mathbf{r}$" refers to voxels intersected by the ray, Δ_i denotes the length of the associated ray-voxel intersection, $\mu(k)$ gives material k's absorption per unit length, and I_0 is the source intensity (obtained from a flat-field calibration scan without the sample). We evaluate these path lengths analytically via 3D-DDA, assuming each voxel is homogeneous—an approach also known as *regular tracking* [Sutton et al. 1999].

Given measured projections $\{I_{\text{ref}}(\mathbf{r})\}$, we define the reconstruction loss as the sum of squared log-differences over all rays:

$$\mathcal{L}(\boldsymbol{X}) = \sum_{\mathbf{r}} \left(\log I(\boldsymbol{X}, \mathbf{r}) - \log I_{\text{ref}}(\mathbf{r})\right)^2. \tag{14}$$

The logarithmic transform is standard in CT reconstruction: it linearizes the exponential Beer-Lambert law into additive optical depths, the form assumed by most reconstruction algorithms. Since both I and I_{ref} are divided by I_0, the source intensity cancels, and the reconstruction operates directly on optical depth $\tau = \sum_{i \in \mathbf{r}} \mu(X_i)\Delta_i$. Each voxel uses a categorical distribution $\boldsymbol{\theta}_i \in \Delta_{m-1}$. The per-voxel expected loss follows Equation 6:

$$\mathcal{L}_i(\boldsymbol{\Theta}, i) = \sum_{k=1}^{m} \theta_{i,k}\, \tilde{\mathcal{L}}_i(k), \tag{15}$$

where $\tilde{\mathcal{L}}_i(k)$ is the approximate loss when X_i is clamped to material k (computed via sampling, mode, or expected-value surrogate as discussed above). This per-voxel loss corresponds to the expected RRE (Section 2.6) under our categorical distribution. Rather than using RRE to evaluate a reconstruction after the fact, our method directly minimizes it. To our knowledge, this is the first optimization approach to do so.

```
mode = "Primal"  # Or "Backward"

# One-ray discrete CT with expectation surrogate
# |μ̂_i = Σ_{k=1}^m θ_{i,k} μ^(k)|,  |τ = Σ_{i∈r} μ̂_i Δ_i|

def eval_ray(r, τ_ref, Θ, {μ^(k)}_{k=1}^m):
    # Pass 1: expected optical depth
    τ = 0
    for (i, Δ_i) in dda_traverse(r):
        μ̂_i = 0
        for k in 1..m:
            μ̂_i += θ_{i,k} * μ^(k)
        τ += μ̂_i * Δ_i

    R = (τ - τ_ref)**2
    if mode == "Primal":
        return R

    # Pass 2: clamped residuals per voxel/material
    g = zeros_like(Θ)
    for (i, Δ_i) in dda_traverse(r):
        μ̂_i = 0
        for k in 1..m:
            μ̂_i += θ_{i,k} * μ^(k)
        τ_base = τ - μ̂_i * Δ_i
        for k in 1..m:
            τ_i = τ_base + μ^(k) * Δ_i
            g[i,k] = (τ_i - τ_ref)**2

    return R, g
```

Listing 1. Efficient two-pass surrogate evaluation: pass one accumulates the expected optical depth, and pass two computes per-voxel per-material residuals without retracing each ray.

Efficient two-pass evaluation. Even with the surrogate, the loss formulation may appear daunting at first: it seems to require averaging many conditional losses and then optimizing their parameters. However, in the absorptive case, the surrogate admits an equivalent reformulation that is local to each projection ray and the intersected set of voxels. In practice, we perform a DDA traversal as in standard CT, though the specifics of the computation differ (Listing 1). To avoid recomputing the conditional loss for every encountered voxel, we use a two-pass scheme that merges the entire derivative evaluation into a single fully fused kernel for efficient GPU processing.

In log space, define the optical depth along a ray under any configuration/approximation as

$$\tau(\boldsymbol{\Theta}, \mathbf{r}) = \sum_{i \in \mathbf{r}} \hat{\mu}_i \Delta_i, \tag{16}$$

where $\hat{\mu}_i$ is either a sampled value $\mu^{(X_i)}$, the mode $\mu^{(\arg\max_k \theta_{i,k})}$, or the expectation $\bar{\mu}_i = \sum_k \theta_{i,k} \mu^{(k)}$. When computing $\hat{\mathcal{L}}_i(k)$, clamping voxel i to material k modifies the optical depth to

$$\tau_i(\boldsymbol{\Theta}, \mathbf{r}; k) = \tau(\boldsymbol{\Theta}, \mathbf{r}) - \hat{\mu}_i \Delta_i + \mu^{(k)} \Delta_i. \tag{17}$$

This local update enables evaluating all m candidate losses per voxel without retracing the ray. We traverse each ray twice via DDA: first to compute $\tau(\boldsymbol{\Theta}, \mathbf{r})$ and the base residual, then to accumulate per-voxel gradients.

3.9 Discrete Computed tomography with scattering

X-ray scattering is typically treated as an artifact in current CT practice. To mitigate its effects, most systems employ anti-scatter grids: thin lead strips between sample and detector, aligned with the source, that absorb scattered photons arriving from off-axis directions. While effective, this approach drastically reduces signal quality and discards potentially useful information.

As Geva et al. [Geva et al. 2018] demonstrate, scattered photons actually carry useful information about tissue composition and can improve reconstruction accuracy. Their work solves an inverse problem for continuous material concentrations; here we address the discrete assignment problem where each voxel must be assigned one of m known materials.

The following experiments with scattering serve as a proof of concept, whose main goal is to show that our framework is compatible with the mathematics of scattering. Scattering in X-ray CT is *complicated*: it involves energy-dependent mixtures of different phenomena (Rayleigh scattering, Compton scattering, the photoelectric effect), some of which depend on the materials' electron shell structure, and energy changes akin to fluorescence for visible light. We are interested in assembling such a comprehensive pipeline, but this is beyond the scope of this work, which focuses on developing a solution to the central challenge of discrete optimization. As a proxy for these scattering effects, we employ the standard Henyey-Greenstein phase function.

Model. Each material $\xi^{(k)}$ in the dictionary is now characterized by absorption and scattering coefficients $\mu_{\mathrm{a}}^{(k)}$ and $\mu_{\mathrm{s}}^{(k)}$ and the phase function $p^{(k)}$. Besides modeling the attenuation of the X-ray source, we further accumulate in-scattered radiance:

$$I(\boldsymbol{\Theta}, \mathbf{r}) = I_{\mathrm{abs}}(\boldsymbol{\Theta}, \mathbf{r}) + \int_0^\infty \hat{\mu}_{\mathrm{s}}(\boldsymbol{\Theta}, \mathbf{r}(t))\, L_s(\mathbf{r}(t))\, \mathrm{d}t, \tag{18}$$

where $\hat{\mu}_{\mathrm{s}}$ is the surrogate scattering coefficient (mode, sample, or expectation) and L_s is the scattered radiance, defined as a spherical integral over the corresponding surrogate phase function $\hat{p}$ and incident radiance L_i:

$$L_s(\boldsymbol{\Theta}, \mathbf{x}, \boldsymbol{\omega}) = \int_{S^2} \hat{p}(\boldsymbol{\Theta}, \mathbf{x}, \boldsymbol{\omega}, \boldsymbol{\omega}')\, L_i(\boldsymbol{\Theta}, \mathbf{x}, \boldsymbol{\omega}')\, \mathrm{d}\boldsymbol{\omega}'. \tag{19}$$

We refer to Pharr et al. [2023, Ch. 11] for a comprehensive review of radiative transfer in volumes.

For each material k, we precompute an expansion of the phase function into zonal harmonics[3]:

$$p^{(k)}(\boldsymbol{\omega}, \boldsymbol{\omega}') \approx \sum_{\ell=0}^{L} p_\ell^{(k)}\, Y_\ell^0(\boldsymbol{\omega} \cdot \boldsymbol{\omega}'). \tag{20}$$

Given spherical harmonics expansions $\mathbf{L}_i$ of incident radiance and $\mathbf{p}$ of the phase function, Equation 19 reduces to an inner product $L_s = \langle \mathbf{L}_i, \mathbf{p} \rangle$. We rotate the phase function expansion into the coordinate frame of $\boldsymbol{\omega}$, which is a cheap operation for zonal harmonics [Jarosz 2008].

We use a two-pass scheme inspired by Moon et al. [2008]. The first pass traces photons from the source, performing random walks that deposit energy into a 3D voxel grid storing spherical harmonics coefficients aligned with the material parameters $\boldsymbol{\Theta}$. The second pass traverses camera rays via DDA, evaluating Equation 18 while accessing both grids in lockstep. Direct illumination from the point source is handled analytically; only indirect incident radiance is stored in the spherical harmonics representation.

[3]The subset of spherical harmonics basis functions with azimuthal symmetry.

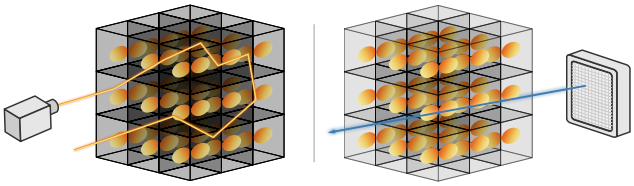


Fig. 6. **Two-Pass Approach for Scattering Volumes**. When reconstructing scattering volumes, we split each optimization step into two passes following Moon et al. [2008]: the first emits photons (orange) from the X-ray source and deposits their contributions to produce a voxelized representation of the incident radiance in a spherical harmonics basis. The second pass casts rays (blue) from the detector and computes the scattered radiance analytically using a product integral with the phase function in the same basis.

As in absorptive CT, we clamp each voxel to candidate materials while other voxels use their approximated values. The precomputed incident radiance field $\mathbf{L}_i$ is held fixed when evaluating candidates, decoupling the per-voxel loss evaluation from global light transport. We acknowledge that this approximation does not propagate derivatives into the photon tracing portion of this two-pass scheme; a fully differentiable scattering implementation remains future work.

Unlike absorptive CT, the system is non-linear: changing a voxel's material affects both its local scattering properties and the transmittance of scattered light from other voxels to the camera.

Geva et al. [2018] optimize continuous element concentrations rather than discrete material assignments, so direct comparison is not appropriate. Instead, we compare against continuous relaxation approaches similar to Gkioulekas et al. [2013] and Nindel et al. [2021], which optimize convex combinations of material properties. To encourage convergence to discrete solutions, we augment the baseline with an entropy regularizer that penalizes diffuse probability distributions.

3.10 Regularization

Discrete CT reconstruction is inherently underdetermined: limited projections and restricted angular coverage mean that multiple material configurations may be consistent with the measured data. Regularization is therefore essential to guide the optimization toward plausible solutions.

Total Variation (TV) regularization [Rudin et al. 1992] is commonly used in CT reconstruction and is particularly well-suited for discrete CT. By suppressing small gradient variations, it promotes solutions that consist of flat regions separated by sharp transitions. Our pipeline optimizes a linear combination of losses $\mathcal{L} + \lambda_{\mathrm{TV}} \mathcal{L}_{\mathrm{TV}}$, where

$$\mathcal{L}_{\mathrm{TV}}(\boldsymbol{\Theta}) = \int_\Omega \left| \nabla \boldsymbol{\Theta}(\mathbf{x}) \right| \mathrm{d}\mathbf{x}. \tag{21}$$

Since each voxel of $\boldsymbol{\Theta}$ is a probability vector rather than a scalar, $|\nabla \boldsymbol{\Theta}|$ denotes the magnitude of the change across all material channels,

which we evaluate using finite differences between neighboring voxels. All methods in our comparison include TV regularization, except for the purely continuous reconstruction (SIRT) and the original DART algorithm, which do not incorporate explicit regularization in their standard formulations.

4 Results

We implemented our method on top of Mitsuba 3 [Jakob et al. 2022b] and the underlying Dr.Jit [Jakob et al. 2022a] just-in-time compiler, using its CUDA backend to run reconstructions on the GPU. Most experiments were conducted on cluster nodes equipped with AMD EPYC 9334 CPUs and NVIDIA L40S GPUs with 48 GiB memory, while some data processing and exploration steps with the real measurements required larger GPU instances (RTX Pro 6000, H100). To enable meaningful runtime comparisons, we implemented all baselines within the same software framework and reused common components where applicable, such as the GPU-based 3D-DDA traversal. For the continuous multilabeling baseline, we used NumPy and the primal-dual hybrid gradient algorithm in the ODL library [Adler et al. 2017] (runtime performance was not relevant for this purely qualitative comparison). We spent approximately 1,800 GPU hours tuning the hyperparameters of the baselines and our method to enable a fair comparison.

4.1 Absorption reconstruction

Figure 7 showcases a matrix of reconstructions of synthetic *phantoms* (i.e., test volumes with known internal structure). This figure shows a subset of a much larger matrix of experiments included in the supplemental material.

To generate challenging test cases, we created complex 2D multi-material patterns using two online pattern generation tools [catchspider2002 2023; Pattern Monster 2023], and converted them into 3D phantoms by rotating the 2D input around the z-axis. These phantoms mimic the geometric complexity found in industrial components such as turbines, heat exchangers with dense flow channels, or honeycomb catalytic converters. The patterns use a 256^2 base resolution and contain 6 (2D) or 7 (3D) materials with evenly spaced absorption coefficients.

We reconstructed 2D scenes with orthographic geometry (512-pixel detector) and 3D scenes with cone-beam geometry (512×512 detector). The leftmost column plots reconstruction performance for all methods as a function of the number of input projections, where the thin gray line at the top indicates 100% accuracy, i.e., correct labeling of all voxels. This visualization reveals how effectively each method exploits the discrete material prior to compensate for limited data. Higher data efficiency directly translates to reduced acquisition time, lower radiation dose for organic specimens, and extended equipment lifetime. The five columns on the right show the ground truth and reconstructions at a representative projection count, chosen as the smallest for which our method produced satisfactory output.

We organize the experiments into several acquisition settings that reflect both theoretical considerations and practical constraints encountered in real-world tomographic imaging. In the *Full 360°* case, all methods receive noise-free projections covering the entire angular domain, sampled using golden ratio angular spacing [Winkelmann et al. 2006]. This idealized setting establishes an upper bound on achievable reconstruction quality and allows us to isolate algorithmic performance from data limitations. Our method quickly reaches perfect reconstruction in this scenario.

The *120°* and *90°* Limited-angle settings address scenarios that are frequently encountered in practice, where physical constraints such as the presence of a gantry, sample holder, or detector housing prevent full rotational access. Such restrictions commonly arise in industrial inspection of large or fixed components, dental and extremity imaging, and electron tomography of thick specimens. Limited-angle tomography is fundamentally more ill-posed than the full-angle case: the missing angular data leaves certain spatial frequencies unobserved, leading to characteristic artifacts such as directional blurring and boundary distortion. Evaluating algorithms under these conditions is essential for assessing their robustness and practical applicability. In these settings, projections are uniformly spaced in angle, and no method achieves perfect accuracy.

The limited-angle results clearly demonstrate the benefit of total variation (TV) regularization in ill-posed reconstruction scenarios. SIRT and DART, which lack such regularization, exhibit substantially degraded performance compared to TVR-DART, which incorporates a TV prior to promote piecewise-constant solutions. Our approach incorporates the same TV prior as TVR-DART, enabling similar robustness to limited data. The accuracy improvements over TVR-DART arise from directly optimizing categorical distributions rather than soft-thresholding a continuous reconstruction.

The final setting, *Poisson*, simulates a high-noise acquisition scenario where the full 360° projection values are corrupted by a Poisson arrival process with $N = 5 \cdot 10^4$ photons per detector element.

Across nearly all settings, our method outperforms the baselines, often doing so by a significant margin.

4.2 Scattering reconstruction

Figure 8 presents an experiment involving a synthetic phantom composed of three scattering materials. This is a particularly challenging setup, as the selected material parameters were chosen to be indistinguishable using an absorption-only simulation (see the left column).

While inverse path tracing methods such as path replay backpropagation [Vicini et al. 2021] could, in principle, be used to address this problem, they typically produce mixed material assignments. Moreover, the high variance of the gradients inherent to such Monte Carlo methods poses significant challenges for optimization. Attempts to regularize these noisy gradients using first-order optimizers are also known to be challenging [Nicolet et al. 2021].

Our method achieves high-fidelity reconstructions in this setting. For this experiment, we used 5 million photons and approximated the indirect radiance using spherical harmonics of order 4. The materials present have Henyey-Greenstein phase functions with anisotropy values in the range $g \in [0.4, 0.5]$.

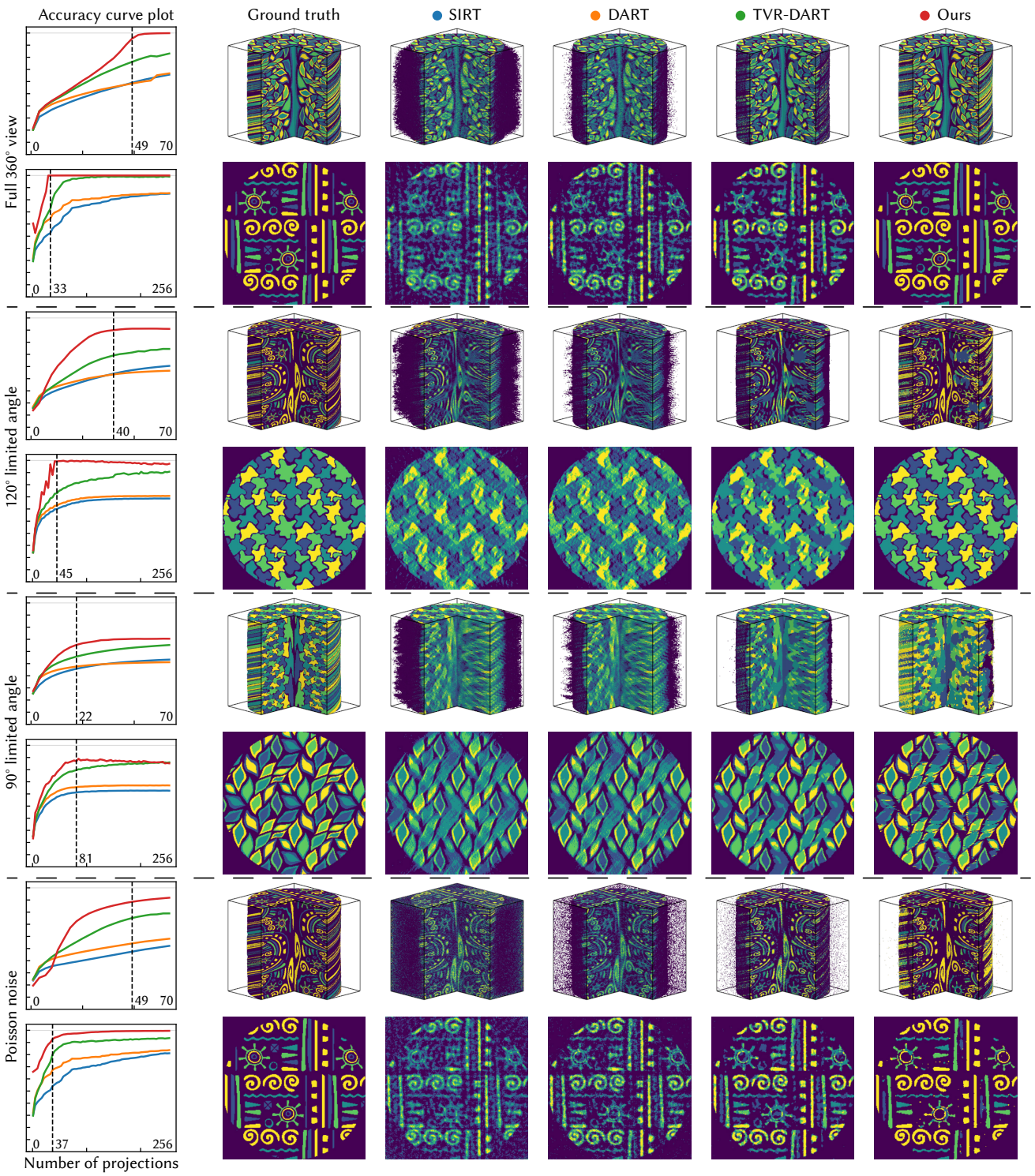


Fig. 7. **Reconstruction of synthetic phantoms (absorption only)**. We evaluate the reconstruction quality on a challenging set of multi-material phantoms. The left column plots accuracy (percentage of correct material assignments) as a function of the number of input projections. The visualizations on the right showcase quality at a specific projection count (indicated by the dashed line), comparing our method to three baselines: SIRT, DART, and TVR-DART. Alternating rows provide 2D (orthographic) and 3D (cone-beam) reconstructions across several scenarios: full 360°, a reduced 120° subset, and a particularly challenging 90° limited-angle case. We also include experiments with Poisson noise added to the full 360° input projections to simulate sensor noise. The supplementary material provides a complete set of experimental results.

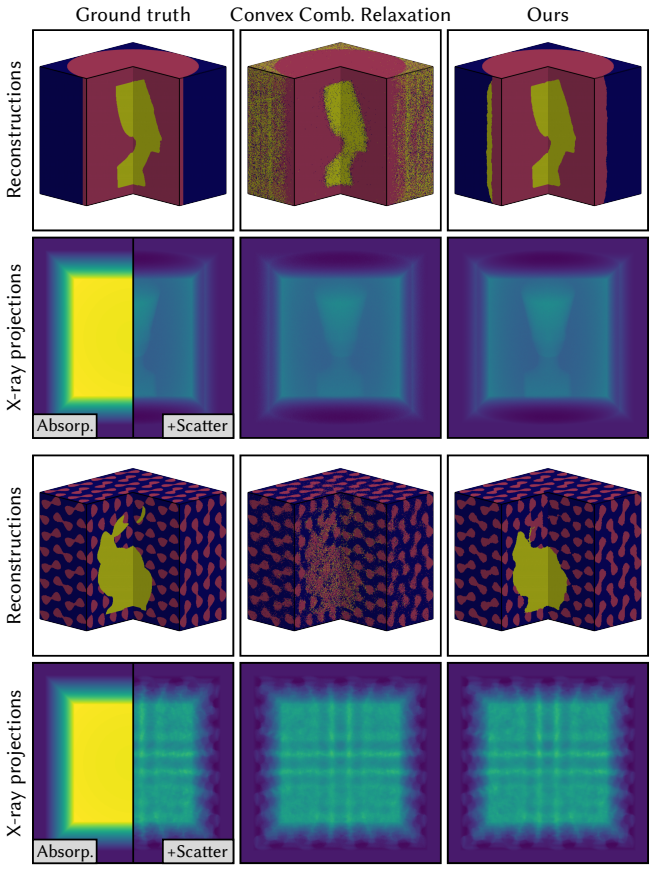


Fig. 8. **Reconstruction with scattering.** All materials in this particularly challenging experiment share the same absorption and differ only in terms of their scattering properties. An absorption-only simulation (left half of ground truth projection) therefore cannot distinguish the materials at all. Even with scattering, it remains difficult to recognize the bunny shape in the projections at the bottom. Our method generates a high-quality segmentation, while standard inverse path tracing via Continuous Convex Relaxation struggles to refine the boundary despite Total Variation regularization.

4.3 Real-world CT reconstruction

We fabricated two multi-material phantoms to also evaluate our method on non-synthetic data.

Phantom design, fabrication, and acquisition. Recent work has demonstrated CT calibration phantoms created using additive manufacturing [Mei et al. 2023; Okkalidis et al. 2024]. Our first step toward multi-material phantoms for discrete CT was to identify printable materials whose X-ray attenuation differs enough to be distinguishable in CT images. We characterized sixteen filaments via CT and selected four with well-separated attenuation profiles: PLA, lightweight PLA (LW-PLA), stone-filled PLA (S-PLA), and ABS.

We fabricated two phantoms targeting different geometric complexities. The first contains triply periodic minimal surface (TPMS) structures [Karakoç 2021] (Gyroid, Schwarz Diamond, Lidinoid, and I-WP), where level-set thresholding of implicit surface functions

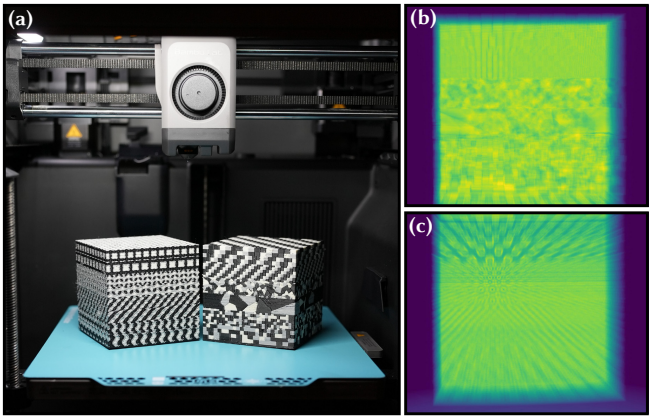


Fig. 9. **FDM 3D printed phantoms.** (a) Two multi-material phantoms fabricated with four materials (PLA, LW-PLA, S-PLA, ABS). CT projections of (b) the stochastic geometry and (c) TPMS phantoms.

creates interleaving material bands. We use varying spatial frequencies to test reconstruction across feature scales. The second contains stochastic geometries (random cube arrangements and Voronoi tessellations with uniform and center-biased seed distributions). All phantoms measure $80 \times 80 \times 80$ mm and were fabricated on a multi-material 3D printer.

We then scanned the phantoms using a ZEISS Metrotom 800 industrial CT system at 130 kV source voltage and 300 µA current. The full acquisition comprises 2880 projections at 1840×1456 pixel resolution, providing a high-quality reference reconstruction against which we evaluate sparse-view performance. We subsample this dataset to create challenging limited-projection scenarios matching those in our synthetic experiments.

3D printing artifacts. Creating multi-material phantoms with FDM 3D printing introduces several practical challenges that make it hard to match ideal discrete reconstructions. The density of printed filament depends on printing parameters such as nozzle temperature, print speed, and how well layers bond together. Additionally, different plastics shrink differently as they cool, which causes geometric distortions and internal stresses, especially when combining materials like ABS and PLA in the same print. These effects create local density variations and artifacts, making it difficult to exactly match the pre-generated ground-truth labels.

Although we selected four filaments with different attenuation profiles, the printing artifacts described above cause their actual attenuation values to vary across the volume. Our selection includes LW-PLA, which uses foaming to create air bubbles that reduce density compared to PLA; ABS, which has a different chemical composition; S-PLA, which contains 50% powdered stone resulting in higher density; and standard PLA. In practice, the actual extinction coefficient varies significantly within each material class due to printing artifacts, and this variation causes the attenuation ranges of different materials to overlap, particularly near boundaries where temperature effects are strongest. As a result, it becomes impossible to distinguish materials by examining pointwise extinction

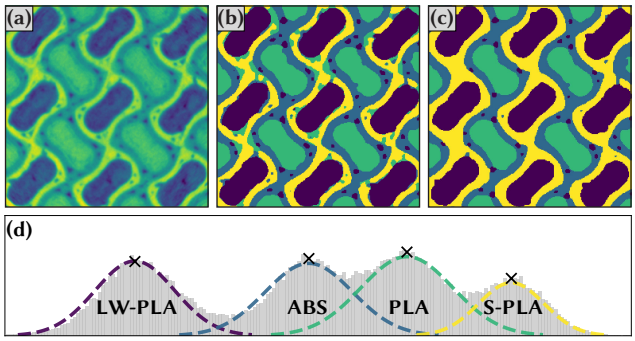


Fig. 10. **Dealing with model inaccuracies.** (a) Continuous reconstruction of a phantom region. (b) Our method with TV regularization assumes constant material attenuation values, but printing imprecisions cause density variations that lead to overlapping attenuation ranges between materials, as shown in the histogram (d). (c) By introducing per-voxel, per-material offsets with a penalty term, our method can absorb these variations while maintaining correct discrete material assignments.

values alone, even with perfectly reconstructed data, as illustrated in Figure 10. To handle these real-world deviations, we extend our framework by allowing a material-dependent continuous offset $\delta_{i,k}$ at each voxel i for each material k, which perturbs the nominal attenuation coefficient via $\mu^{(k)} \mapsto \mu^{(k)} + \delta_{i,k}$. We penalize large offsets using a Gaussian prior with per-material strength:

$$\mathcal{L}_{\text{offset}} = \lambda_\delta \sum_i \sum_{k=1}^{m} \frac{\delta_{i,k}^2}{2\sigma_{\delta,k}^2},$$

where each $\sigma_{\delta,k}$ controls how much deviation from material k's ideal attenuation is allowed. This lets the optimization absorb small density variations and measurement errors while still selecting discrete material assignments, as shown in Figure 10.

Reconstructions of the real phantoms. Figure 12 presents reconstruction results for our two 3D-printed multi-material phantoms.

Due to the printing artifacts discussed in the previous section, no pixel-accurate ground truth is available. We therefore use a high-quality FDK reconstruction from all 2880 projections as a reference for visual comparison. This continuous reconstruction reveals the internal structure well, though it cannot serve as a discrete ground truth due to the inherent attenuation variations within each material class.

We evaluate all methods under sparse-view conditions: 64 projections for the stochastic geometry phantom and 128 projections for the TPMS phantom.

SIRT, lacking any discrete prior, produces reconstructions with poorly defined material boundaries under these sparse-view conditions. While it occasionally captures the extreme materials correctly (since we clamp the reconstruction to the valid attenuation range), the overall segmentation quality is limited by the ill-posed nature of the reconstruction problem. DART improves upon SIRT by incorporating a discrete prior and boundary smoothing, which helps stabilize the reconstruction and avoid such extreme assignments.

TVR-DART achieves smoother results by combining soft segmentation with total variation regularization. However, its formulation encounters difficulties at boundaries between materials with non-adjacent attenuation values. The method uses a monotonic soft segmentation function that maps continuous attenuation to discrete gray levels along a single axis. When two materials with distant attenuation values meet directly (e.g., LW-PLA and S-PLA), the TV regularization can favor distributing the gradient over multiple smaller steps rather than one large transition. This manifests as characteristic rings of intermediate materials at such boundaries, visible in the reconstructions.

Our method avoids these artifacts through its categorical formulation. The per-material offset parameters absorb local density variations without blurring the material boundaries. This allows materials with overlapping or non-adjacent attenuation ranges to share boundaries without introducing intermediate material bands, resulting in the crisp transitions visible in Figure 12.

Limited-angle reconstructions. Figure 13 extends this evaluation to limited-angle acquisitions with 90° and 120° angular ranges, keeping the same projection counts as in the full-360° case (64 for the stochastic geometry phantom, 128 for the TPMS phantom). The limited angle leaves certain feature orientations under-constrained: at 90°, only structures with a particular orientation are recovered while their rotated counterparts are largely averaged away in all reconstructions. The additional 30° available at 120° reduces this asymmetry: ellipsoidal features in the TPMS phantom that are averaged away at 90° (row 2) reappear at 120° (row 5). All methods degrade compared to the full-360° results in Figure 12, as expected. Even so, our method preserves the qualitative behavior of the full-view setting: material interfaces stay sharp and transitions between non-adjacent attenuation values are recovered cleanly, without the intermediate-material rings discussed above.

4.4 Hyperparameters

All evaluated methods depend on hyperparameters, some of which require careful tuning to achieve good performance. We used the tree-structured Parzen sampler (`TPESampler`) from the Optuna [Akiba et al. 2019] library to optimize them on a challenging limited-angle phantom, using the resulting configurations consistently across experiments.

The hyperparameters for each method are:

- **SIRT**: This method has a single parameter that controls the number of forward/back-projection iterations. 20 iterations are common in the literature, with 200 mentioned as an upper bound for extreme cases. We generally use 200, and 1000 in Figure 1 and Figure 12. However, reconstruction quality typically plateaus early, and higher iteration counts offer little additional benefit. To convert the continuous reconstruction into a discrete one, we threshold the result at the midpoint of the absorption intervals.
- **DART**: We initialize with 200 iterations of SIRT, then perform 5 additional SIRT steps per DART iteration. We set the free pixel probability to 15% following the authors' recommendation.
- **TVR-DART**: This method uses a quasi-second-order optimizer, specifically L-BFGS [Liu and Nocedal 1989]. We set its history size to 5, as larger values exceed the available 48 GiB of GPU memory.

We use a total variation weight of $\lambda = 10^{-3}$ using a Huber loss with $\varepsilon = 10^{-4}$, and a soft thresholding parameter of $K = 8$. This method converges significantly slower than the others, so we cap the number of iterations at 10,000. For the real data experiments, we found that $K = 0.1$ and $\lambda = 10^{-1}$ produce the best results.

- **Ours**: We use the Adam optimizer [Kingma and Ba 2014] with a learning rate of $\eta = 2 \cdot 10^{-2}$, and set the L_2 total variation weight to $\lambda = 10^{-4}$. For the real data experiments, we use $\lambda = 10^{-3}$ and $\lambda_\delta = 0.1$; the per-material offset standard deviations $\sigma_{\delta,k}$ are listed in Table 1 (Section 3 in the supplemental material).

4.5 Performance data

Figure 11 compares reconstruction accuracy against wall-clock time for the 120° synthetic limited-angle experiments. All methods share a similar computational structure, traversing the voxel grid twice per iteration for forward projection and gradient computation. Our method has some overhead from evaluating all K candidate materials per voxel, but this broader exploration of the solution space leads to faster convergence in accuracy per unit time.

SIRT and DART plateau quickly but show little improvement thereafter; SIRT is limited by its continuous relaxation, while DART's hard thresholding can lock in suboptimal assignments early. TVR-DART achieves higher accuracy but converges slowly; while quasi-Newton methods like L-BFGS can converge faster than first-order optimizers, they require line searches, and their memory overhead forced us to limit the history size, reducing their effectiveness. Our use of the modified natural gradient with Adam helps here, allowing larger step sizes while remaining stable near convergence.

5 Conclusion

Optimization over large discrete spaces normally involves a difficult choice: accept the inefficiency of stochastic methods like simulated annealing, or relax the problem and sacrifice solution quality. The key problem is the absence of a downhill direction—there is only *elsewhere*, in a combinatorially vast space. Our method combines several insights to resolve this quandary: rather than optimizing a reconstruction, it refines statistical belief. It cannot search the full space, but it can exhaustively probe the neighborhood of any given state, and this proves sufficient. The distribution adapts as local exploration reveals promising directions and eventually concentrates into a single choice.

The scattering-aware variant of our method makes several approximations: we simplified the X-ray physics by ignoring energy and material-dependent directional scattering. We also hold the incident radiance field fixed while evaluating gradients. This works for typical X-ray CT settings, where scattering is directly observed, but it may be restrictive outside tomography.

Discrete optimization problems arise throughout graphics, from selecting among BRDF models to choosing procedural noise seeds. Global illumination complicates this by coupling parameters through interreflection. Extending our approach with efficient unbiased Monte Carlo gradient estimates for such settings is a natural direction for future work.

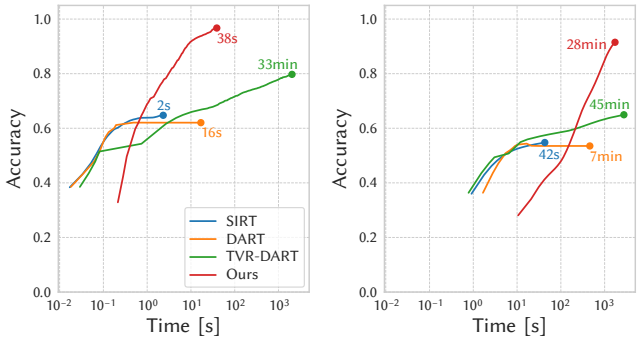


Fig. 11. **Reconstruction accuracy vs. time**. Comparison of reconstruction accuracy and speed for the 120° synthetic limited-angle experiment (see Fig. 7) for both 2D (left, 45 projections) and 3D (right, 40 projections). SIRT and DART converge quickly but plateau at relatively low accuracy, with little improvement even when increasing the number of iterations. Our method reaches a higher accuracy within a significantly shorter computation time.

References

Jonas Adler, Holger Kohr, and Ozan Öktem. 2017. Operator discretization library (ODL). *Zenodo* (2017).

Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. 2019. Optuna: A Next-generation Hyperparameter Optimization Framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

Shun-ichi Amari. 2016. *Information Geometry and Its Applications* (1st ed.). Springer Publishing Company, Incorporated.

A.H. Andersen and A.C. Kak. 1984. Simultaneous Algebraic Reconstruction Technique (SART): A superior implementation of the ART algorithm. *Ultrasonic Imaging* 6, 1 (1984), 81–94.

Kees Joost Batenburg and Jan Sijbers. 2011. DART: A Practical Reconstruction Algorithm for Discrete Tomography. *IEEE Transactions on Image Processing* 20, 9 (2011).

Christopher M Bishop and Nasser M Nasrabadi. 2006. *Pattern recognition and machine learning*. Vol. 4. Springer.

F. Bleichrodt, F. Tabak, and K.J. Batenburg. 2014. SDART: An algorithm for discrete tomography from noisy projections. *Computer Vision and Image Understanding* 129 (2014). Special section: Advances in Discrete Geometry for Computer Imagery.

R N Bracewell and A C Riddle. 1967. Inversion of fan-bam scans in radio astronomy. *Astrophys. J., 150: 427-34(Nov. 1967).* (01 1967).

catchspider2002. 2023. svelte-svg-patterns: Create seamless patterns in SVG. https://github.com/catchspider2002/svelte-svg-patterns GitHub repository for SVG pattern generation.

A. M. Cormack. 1964. Representation of a Function by Its Line Integrals, with Some Radiological Applications. II. *Journal of Applied Physics* 35, 10 (10 1964), 2908–2913.

Andrei Dabravolski, Kees Joost Batenburg, and Jan Sijbers. 2014. A Multiresolution Approach to Discrete Tomography Using DART. *PLOS ONE* 9, 9 (09 2014), 1–10.

Samuel Daulton, Xingchen Wan, David Eriksson, Maximilian Balandat, Michael A Osborne, and Eytan Bakshy. 2022. Bayesian optimization over discrete and mixed spaces via probabilistic reparameterization. *Advances in Neural Information Processing Systems* 35 (2022), 12760–12774.

L. A. Feldkamp, L. C. Davis, and J. W. Kress. 1984. Practical cone-beam algorithm. *J. Opt. Soc. Am. A* 1, 6 (Jun 1984), 612–619.

Daniel Frenkel, Nathanaël Six, Jan De Beenhouwer, and Jan Sijbers. 2023. Tabu-DART: a dynamic update strategy for efficient discrete algebraic reconstruction. *The Visual Computer* 39, 10 (01 Oct 2023), 4671–4683.

R.J. Gardner, P. Gritzmann, and D. Prangenberg. 1999. On the computational complexity of reconstructing lattice sets from their X-rays. *Discrete Mathematics* 202, 1 (1999).

Yan Gerard. 2009. About the complexity of timetables and 3-dimensional discrete tomography: A short proof of NP-hardness. In *Combinatorial Image Analysis: 13th International Workshop, IWCIA 2009, Playa del Carmen, Mexico, November 24-27, 2009. Proceedings 13*. Springer, 289–301.

Adam Geva, Yoav Y Schechner, Yonatan Chernyak, and Rajiv Gupta. 2018. X-ray computed tomography through scatter. In *Proceedings of The European Conference on Computer Vision (ECCV)*. 34–50.

Peter Gilbert. 1972. Iterative methods for the three-dimensional reconstruction of an object from projections. *Journal of Theoretical Biology* 36, 1 (1972), 105–117.

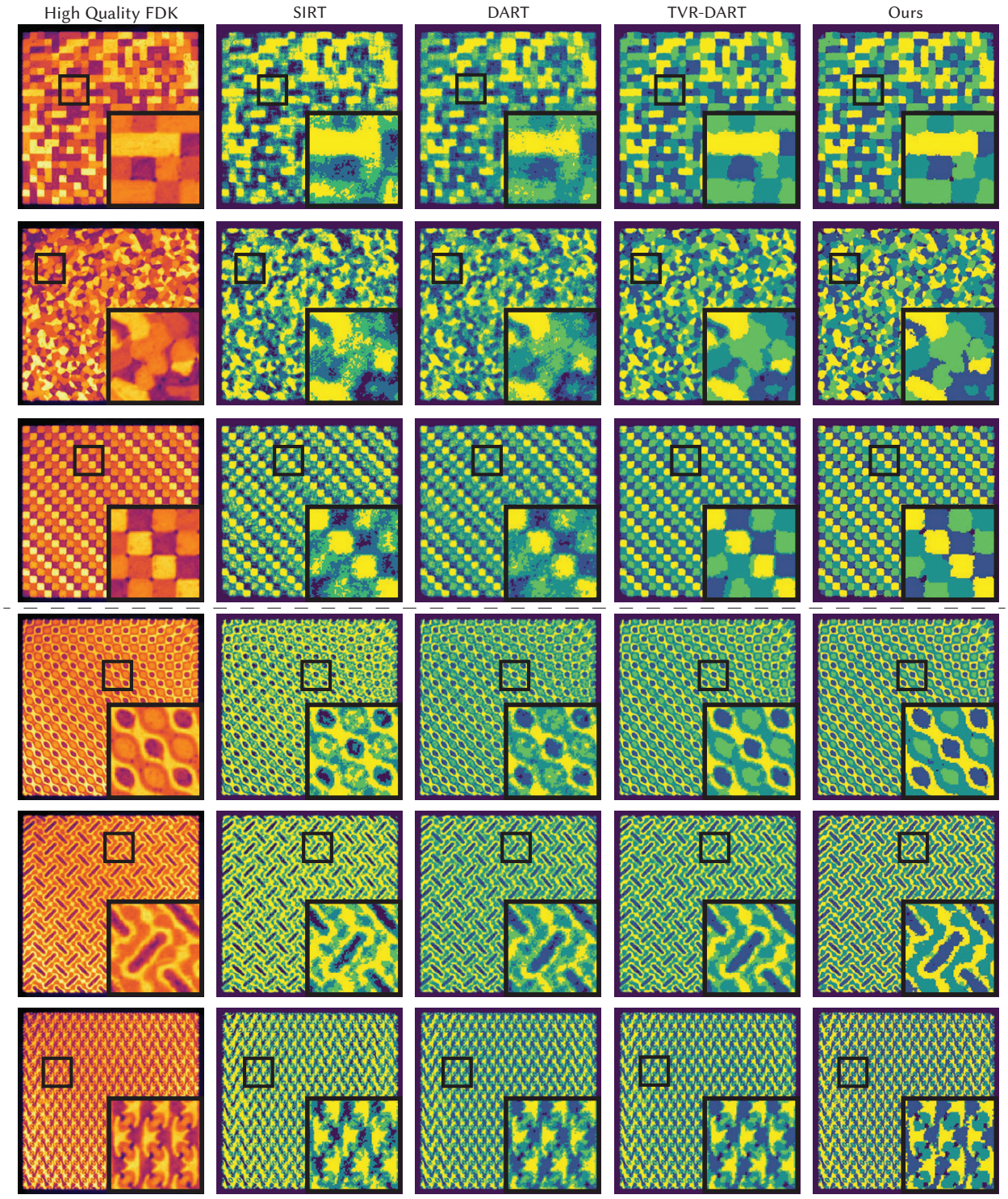


Fig. 12. **Reconstruction of 3D printed phantoms**. We evaluate reconstruction quality on two multi-material phantoms fabricated via FDM printing with four materials (PLA, LW-PLA, S-PLA, ABS). The top three rows show slices from a stochastic geometry phantom; the bottom three rows show a TPMS phantom. The leftmost column is a high-quality FDK reference obtained from all 2880 projections. It is continuous and therefore uses a different color map. The remaining columns compare our method against three baselines using 64 views (top three rows) and 128 views (bottom three rows), all with full 360° angular coverage at 512^3 resolution. Slices are selected to highlight challenging regions. Our method produces sharper material interfaces compared to the baselines.

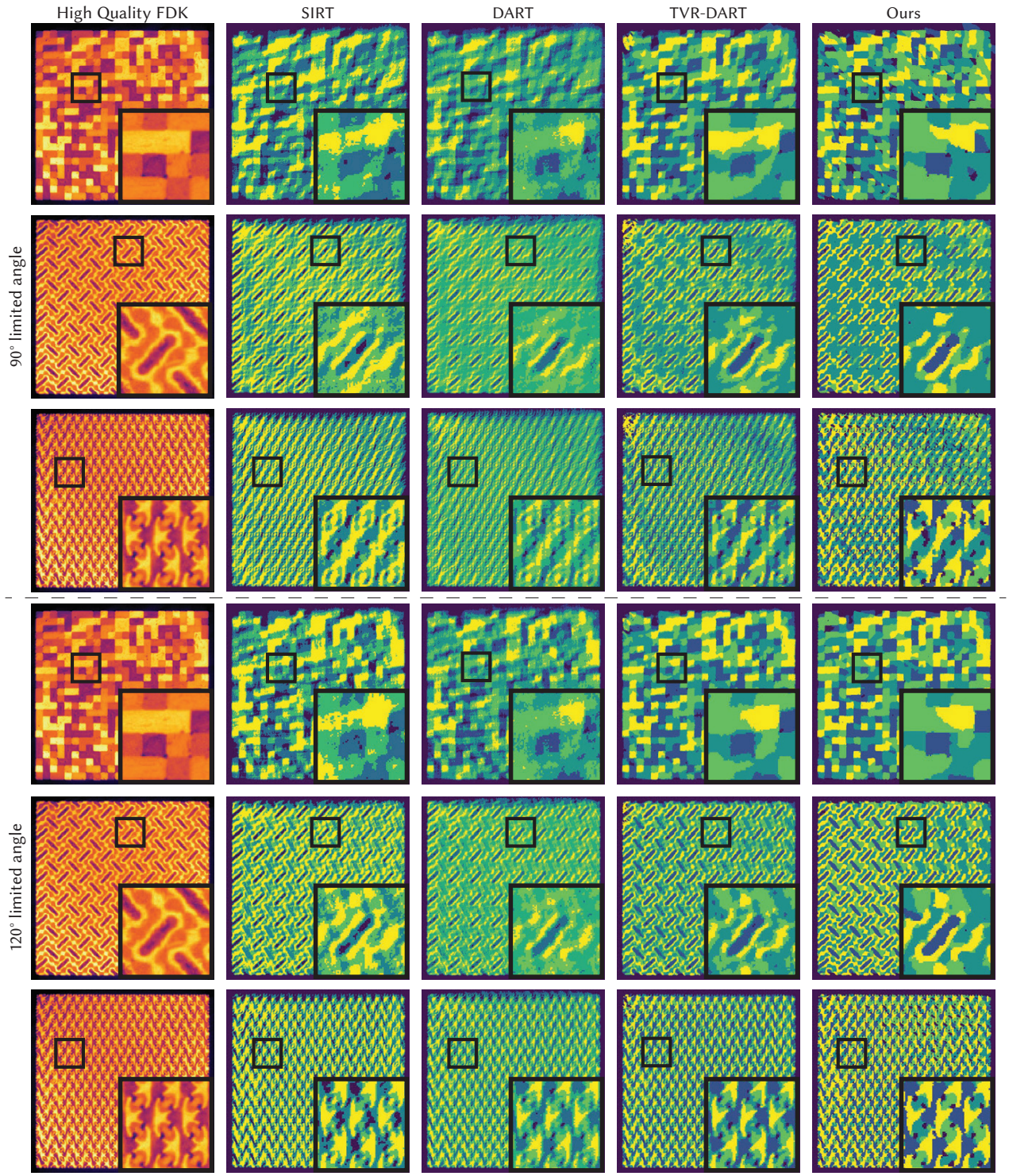


Fig. 13. **Reconstruction of 3D printed phantoms under limited angular coverage.** We restrict the acquisition to a 90° (top three rows) or 120° (bottom three rows) angular range. Within each block, the first row shows a slice from the stochastic geometry phantom (64 projections) and the following two rows show slices from the TPMS phantom (128 projections). Projection counts and slice locations match Figure 12 for direct comparison with the full-360° case. All reconstructions are at 512^3 resolution. Our method retains the sharp material interfaces seen in the full-view setting, while the baselines produce visibly less sharp boundaries.

Ioannis Gkioulekas, Shuang Zhao, Kavita Bala, Todd Zickler, and Anat Levin. 2013. Inverse volume rendering with material dictionaries. *ACM Trans. Graph.* 32, 6, Article 162 (Nov. 2013), 13 pages. doi:10.1145/2508363.2508377

James Gregson, Michael Krimerman, Matthias B. Hullin, and Wolfgang Heidrich. 2012. Stochastic tomography and its applications in 3D imaging of mixing fluids. *ACM Trans. Graph.* 31, 4, Article 52 (July 2012), 10 pages. doi:10.1145/2185520.2185548

Yutao He, Wenquan Ming, Ruohan Shen, and Jianghua Chen. 2022. IDART: An Improved Discrete Tomography Algorithm for Reconstructing Images With Multiple Gray Levels. *IEEE Transactions on Image Processing* 31 (2022), 2608–2619.

Kuidong Huang, Zongzhao Gao, Fuqiang Yang, Hua Zhang, and Dinghua Zhang. 2023. An improved discrete algebraic reconstruction technique for limited-view based on gray mean value guidance. *Journal of Nondestructive Evaluation* 42, 1 (2023), 6.

Wenzel Jakob, Sébastien Speierer, Nicolas Roussel, Merlin Nimier-David, Delio Vicini, Tizian Zeltner, Baptiste Nicolet, Miguel Crespo, Vincent Leroy, and Ziyi Zhang. 2022b. *Mitsuba 3 renderer.* https://mitsuba-renderer.org.

Wenzel Jakob, Sébastien Speierer, Nicolas Roussel, and Delio Vicini. 2022a. Dr.Jit: A Just-In-Time Compiler for Differentiable Rendering. *Transactions on Graphics (Proceedings of SIGGRAPH)* 41, 4 (July 2022).

Eric Jang, Shixiang Gu, and Ben Poole. 2017. Categorical Reparameterization with Gumbel-Softmax. arXiv:1611.01144 [stat.ML] https://arxiv.org/abs/1611.01144

Wojciech Jarosz. 2008. *Efficient Monte Carlo Methods for Light Transport in Scattering Media.* Ph. D. Dissertation. UC San Diego.

Alp Karakoç. 2021. RegionTPMS — Region based triply periodic minimal surfaces (TPMS) for 3-D printed multiphase bone scaffolds with exact porosity values. *SoftwareX* 16 (2021), 100835. doi:10.1016/j.softx.2021.100835

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 2023. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Transactions on Graphics* 42, 4 (July 2023).

Diederik P. Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. *CoRR* abs/1412.6980 (2014). https://api.semanticscholar.org/CorpusID:6628106

Dong C. Liu and Jorge Nocedal. 1989. On the limited memory BFGS method for large scale optimization. *Math. Program.* 45, 1–3 (Aug. 1989), 503–528.

F. Javier Maestre-Deusto, Giovanni Scavello, Joaquín Pizarro, and Pedro L. Galindo. 2011. ADART: An Adaptive Algebraic Reconstruction Algorithm for Discrete Tomography. *IEEE Transactions on Image Processing* 20, 8 (2011), 2146–2152.

Kai Mei, Pouyan Pasyar, Michael Geagan, Leening P Liu, Nadav Shapira, Grace J Gang, J Webster Stayman, and Peter B Noël. 2023. Design and fabrication of 3D-printed patient-specific soft tissue and bone phantoms for CT imaging. *Scientific reports* 13, 1 (2023), 17495.

Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. 2020. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In *ECCV*.

Jonathan T. Moon, Bruce Walter, and Steve Marschner. 2008. Efficient multiple scattering in hair using spherical harmonics. In *ACM SIGGRAPH 2008 Papers* (Los Angeles, California) *(SIGGRAPH '08)*. Association for Computing Machinery, New York, NY, USA, Article 31.

Baptiste Nicolet, Alec Jacobson, and Wenzel Jakob. 2021. Large steps in inverse rendering of geometry. *ACM Trans. Graph.* 40, 6, Article 248 (Dec. 2021).

Thomas Klaus Nindel, Tomáš Iser, Tobias Rittig, Alexander Wilkie, and Jaroslav Křivánek. 2021. A gradient-based framework for 3D print appearance optimization. *ACM Trans. Graph.* 40, 4, Article 178 (July 2021), 15 pages. doi:10.1145/3450626.3459844

Filippos Okkalidis, Chrysoula Chatzigeorgiou, Nikiforos Okkalidis, Nikolay Dukov, Minko Milev, Zhivko Bliznakov, Giovanni Mettivier, Paolo Russo, and Kristina Bliznakova. 2024. Characterization of commercial and custom-made printing filament materials for Computed Tomography imaging of radiological phantoms. *Technologies* 12, 8 (2024), 139.

Pattern Monster. 2023. Pattern Monster - SVG Pattern Generator. https://pattern.monster Online SVG pattern generator tool.

Matt Pharr, Wenzel Jakob, and Greg Humphreys. 2023. *Physically based rendering: From theory to implementation.* MIT Press.

J. Radon. 1917. Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten. *Akad. Wiss.* 69 (1917), 262–277.

G. N. Ramachandran and A. V. Lakshminarayanan. 1971. Three-dimensional Reconstruction from Radiographs and Electron Micrographs: Application of Convolutions instead of Fourier Transforms. *Proceedings of the National Academy of Sciences* 68, 9 (1971), 2236–2240.

Essam A Rashed and Hiroyuki Kudo. 2012. Statistical image reconstruction from limited projection data with intensity priors. *Physics in Medicine & Biology* 57, 7 (2012).

Tom Roelandts, K. Joost Batenburg, Arnold J. den Dekker, and Jan Sijbers. 2014. The reconstructed residual error: A novel segmentation evaluation measure for reconstructed images in tomography. *Computer Vision and Image Understanding* 126 (2014), 28–37.

Darius Rückert, Yuanhao Wang, Rui Li, Ramzi Idoughi, and Wolfgang Heidrich. 2022. NeAT: neural adaptive tomography. *ACM Trans. Graph.* 41, 4, Article 55 (July 2022).

Leonid I Rudin, Stanley Osher, and Emad Fatemi. 1992. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena* 60, 1-4 (1992), 259–268.

Larry A. Shepp and Yehuda Vardi. 1983. Maximum Likelihood Reconstruction for Emission Tomography. *IEEE Transactions on Medical Imaging* 1 (1983), 113–122.

Nathanaël Six, Jan De Beenhouwer, and Jan Sijbers. 2019. poly-DART: A discrete algebraic reconstruction technique for polychromatic X-ray CT. *Opt. Express* 27, 23 (Nov 2019), 33670–33682.

Yu Sun, Jiaming Liu, Mingyang Xie, Brendt Wohlberg, and Ulugbek S. Kamilov. 2021. CoIL: Coordinate-Based Internal Learning for Tomographic Imaging. *IEEE Transactions on Computational Imaging* 7 (2021), 1400–1412.

TM Sutton, FB Brown, FG Bischoff, DB MacMillan, CL Ellis, JT Ward, CT Ballinger, DJ Kelly, and L Schindler. 1999. *The physical models and statistical procedures used in the RACER Monte Carlo code.* Technical Report. Knolls Atomic Power Lab.(KAPL), Niskayuna, NY (United States); Albany.

Wim van Aarle, Kees Joost Batenburg, and Jan Sijbers. 2012. Automatic Parameter Estimation for the Discrete Algebraic Reconstruction Technique (DART). *IEEE Transactions on Image Processing* 21, 11 (2012), 4608–4621.

Delio Vicini, Sébastien Speierer, and Wenzel Jakob. 2021. Path Replay Backpropagation: Differentiating Light Paths using Constant Memory and Linear Time. *Transactions on Graphics (Proceedings of SIGGRAPH)* 40, 4 (Aug. 2021), 108:1–108:14. doi:10.1145/3450626.3459804

Feiran Wang, Jiachen Tao, Junyi Wu, Haoxuan Wang, Bin Duan, Kai Wang, Zongxin Yang, and Yan Yan. 2025. X-Field: A Physically Grounded Representation for 3D X-ray Reconstruction. *arXiv preprint arXiv:2503.08596* (2025).

Weiran Wang and Miguel Á. Carreira-Perpiñán. 2013. Projection onto the probability simplex: An efficient algorithm with a simple proof, and an application. arXiv:1309.1541 [cs.LG]

Stefanie Winkelmann, Tobias Schaeffter, Thomas Koehler, Holger Eggers, and Olaf Doessel. 2006. An optimal radial profile order based on the Golden Ratio for time-resolved MRI. *IEEE transactions on medical imaging* 26, 1 (2006), 68–76.

Guangming Zang, Ramzi Idoughi, Rui Li, Peter Wonka, and Wolfgang Heidrich. 2021. IntraTomo: Self-supervised Learning-based Tomography via Sinogram Synthesis and Prediction. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)* (2021), 1940–1950.

Ruyi Zha, Tao Jun Lin, Yuanhao Cai, Jiwen Cao, Yanhao Zhang, and Hongdong Li. 2024. R^2-Gaussian: Rectifying Radiative Gaussian Splatting for Tomographic Reconstruction. In *Advances in Neural Information Processing Systems (NeurIPS)*.

Ziyi Zhang, Nicolas Roussel, and Wenzel Jakob. 2024. Many-Worlds Inverse Rendering. arXiv:2408.16005 [cs.CV]

Ziyi Zhang, Nicolas Roussel, Thomas Müller, Tizian Zeltner, Merlin Nimier-David, Fabrice Rousselle, and Wenzel Jakob. 2025. Radiance Surfaces: Optimizing Surface Representations with a 5D Radiance Field Loss. In *Proceedings of the ACM SIGGRAPH Conference Papers (SIGGRAPH Conference Papers '25)* (Vancouver, BC, Canada). Association for Computing Machinery, New York, NY, USA, 15 pages. doi:10.1145/3721238.3730713

Xiaodong Zhuge, Willem Jan Palenstijn, and Kees Joost Batenburg. 2016. TVR-DART: A More Robust Algorithm for Discrete Tomography From Limited Projection Data With Automated Gray Value Estimation. *IEEE Transactions on Image Processing* 25, 1 (2016), 455–468.

Matthias Zisler, Stefania Petra, Claudius Schnörr, and Christoph Schnörr. 2016. Discrete Tomography by Continuous Multilabeling Subject to Projection Constraints. In *Pattern Recognition*, Bodo Rosenhahn and Bjoern Andres (Eds.). Springer International Publishing, Cham, 261–272.